

Adjusted Priors for Bayes Factors Involving Reparameterized Order Constraints

Daniel W. Heck

University of Mannheim

Eric-Jan Wagenmakers

University of Amsterdam

Author Note

Daniel W. Heck, School of Social Sciences, Department of Psychology, University of Mannheim, Germany. Eric-Jan Wagenmakers, University of Amsterdam, Department of Psychological Methods, Amsterdam, the Netherlands.

This work was supported in part by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation. Code and data associated with this paper have been lodged with the Open Science Framework at <https://osf.io/cz827/>.

Correspondence concerning this article should be addressed to Daniel W. Heck, Department of Psychology, School of Social Sciences, University of Mannheim, Schloss EO 254, D-68131 Mannheim, Germany. E-mail: dheck@mail.uni-mannheim.de; Phone: (+49) 621 181 2145.

Abstract

Many psychological theories that are instantiated as statistical models imply order constraints on the model parameters. To fit and test such restrictions, order constraints of the form $\theta_i \leq \theta_j$ can be reparameterized with auxiliary parameters $\eta \in [0, 1]$ to replace the original parameters by $\theta_i = \eta \cdot \theta_j$. This approach is especially common in multinomial processing tree (MPT) modeling because the reparameterized, less complex model also belongs to the MPT class. Here, we discuss the importance of adjusting the prior distributions for the auxiliary parameters of a reparameterized model. This adjustment is important for computing the Bayes factor, a model selection criterion that measures the evidence in favor of an order constraint by trading off model fit and complexity. We show that uniform priors for the auxiliary parameters result in a Bayes factor that differs from the one that is obtained using a multivariate uniform prior on the order-constrained original parameters. As a remedy, we derive the adjusted priors for the auxiliary parameters of the reparameterized model. The practical relevance of the problem is underscored with a concrete example using the multi-trial pair-clustering model.

Keywords: Inequality constraints, model selection, MPT modeling, model complexity, encompassing prior.

Adjusted Priors for Bayes Factors Involving Reparameterized Order Constraints

Linear order constraints of the form $\theta_i \leq \theta_j$ are important for many substantive hypotheses in mathematical psychology (Hojtink, 2011; Iverson, 2006). In the present paper, we are mainly concerned with order constraints in multinomial processing tree (MPT) models (Batchelder and Riefer, 1999; Erdfelder et al., 2009), a class of cognitive models that accounts for response frequencies by a finite number of underlying processes. In MPT models, the parameters θ represent conditional and unconditional probabilities that specific cognitive states are entered. Hence, order constraints are necessary to test whether some of the underlying processes occur more often under specific conditions (Knapp and Batchelder, 2004).

Linear order constraints on MPT parameters can easily be represented using auxiliary parameters, resulting in a new, reparameterized MPT model. Importantly, such an order-constrained model makes more specific predictions than the original, unconstrained model and is less complex even though both models have the same number of free parameters (Myung, 2000; Vanpaemel, 2009). Such a difference in complexity is taken into account by the Bayes factor, the standard Bayesian model selection tool that is able to quantify the evidence in favor of the order constraint (Hojtink, 2011; Klugkist, Laudy, and Hoijtink, 2005).

In the present paper, we show that the combination of both methods –Bayes factors for order-constraints in a reparameterized MPT model– is prone to an inadvertent choice of prior distributions that do not match the actual prior beliefs. Specifically, we illustrate the substantial difference between results obtained under a uniform prior on the original, constrained parameter space and those obtained under a uniform prior on the auxiliary parameters. Whereas the former meets the intuition that all admissible, original parameter vectors are equally likely, the latter leads to an overweighting of small values of the original parameters. We derive adjusted priors for the auxiliary parameters that imply a uniform prior on the original parameters and underscore the importance of this adjustment in case of the pair-clustering model (Batchelder and Riefer, 1980). Despite our focus on MPT models, the present paper

highlights the importance of carefully choosing priors for reparameterized models in general.

Reparameterization of Order Constraints in MPT Models

When maximizing the likelihood of a model, many optimization procedures require that each parameter varies independently on a fixed range (e.g., between zero and one). However, if a model contains order constraints of the form $\theta_i \leq \theta_j$, the admissible values for one parameter depend on the values of the other parameters. As a practical solution, models are often reparameterized into statistically equivalent models with independent parameters. For instance, within a bounded two-dimensional parameter space $\boldsymbol{\theta} \in [a, b] \times [a, b]$, the constraint $\theta_1 \leq \theta_2$ can be implemented by auxiliary parameters defined as $\eta_2 = \theta_2$ and $\eta_1 = \theta_1/\theta_2$ (Knapp and Batchelder, 2004). The equivalent expression $\theta_1 = \eta_1\theta_2$ provides an intuitive interpretation of η_1 as a ‘decrease parameter’ between zero and one. The reparameterized model with two independent parameters $(\eta_1, \eta_2) \in [0, 1] \times [a, b]$ is statistically equivalent to the original model because both models imply the same set of probability distributions over the data.

MPT models are special with regard to linear order constraints, because the reparameterization results in a new, less flexible MPT model (Klauer, Singmann, and Kellen, 2015; Knapp and Batchelder, 2004). In other words, the use of auxiliary parameters $\eta \in [0, 1]$ to implement order constraints is equivalent to adding one or more branches in the graphical representation of a processing tree. This result facilitates testing order constraints with standard software for MPT models (e.g., Moshagen, 2010; Singmann and Kellen, 2013) or hierarchical extensions (e.g., Klauer, 2010; Smith and Batchelder, 2010). Given these benefits, it is not surprising to find many applications of this approach, for instance, concerning pair-clustering in memory (Bröder et al., 2008; Riefer et al., 2002), outcome-based strategy selection in decision making (Hilbig and Moshagen, 2014), recognition memory (Kellen and Klauer, 2015), and stochastic dominance of binned response time distributions (Heck and Erdfelder, in press).

The Bayes Factor

Bayesian model selection is based on the Bayes factor (Jeffreys, 1961; Kass and Raftery, 1995), defined as the odds of the data under models \mathcal{M}_0 vs. \mathcal{M}_1 :

$$B_{01} = \frac{p(\mathbf{y} \mid \mathcal{M}_0)}{p(\mathbf{y} \mid \mathcal{M}_1)}. \quad (1)$$

In order to obtain the marginal likelihood of the data \mathbf{y} given the model \mathcal{M}_0 , the likelihood f_0 is integrated over the parameter space Θ weighted by the prior π_0 ,

$$p(\mathbf{y} \mid \mathcal{M}_0) = \int_{\Theta} f_0(\mathbf{y} \mid \boldsymbol{\theta}) \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (2)$$

Mathematically, this is the prior-weighted average of the likelihood (Wagenmakers et al., 2015), which explains the substantial impact of the prior on model selection. The Bayes factor can directly be interpreted as the evidence in favor of one model compared to another and is ideally suited to evaluate order constraints, which restrict the volume of the parameter space Θ (Klugkist and Hoijtink, 2007; Klugkist, Laudy, and Hoijtink, 2005). If an order constraint restricts the parameters to be in the admissible subspace $\Theta' \subset \Theta$, the integral in Eq. 2 will be large if the weighted likelihood on the subspace Θ' is high, but small if the weighted likelihood on the subspace Θ' is low.

When testing order constraints on bounded parameters, an important and common prior is the uniform distribution, which has a constant density on the restricted parameter space Θ' and puts equal probability mass on all admissible parameter vectors (Klugkist, Laudy, and Hoijtink, 2005). Note that this prior is theoretically appropriate to model, for instance, equally-likely error probabilities (Lee, 2016; Myung, Karabatsos, and Iverson, 2005), but might be inappropriate for other applications (e.g., guessing probabilities expected to be around 50%). Importantly, this multivariate uniform distribution refers to the original parameters $\boldsymbol{\theta}$. If order constraints are implemented by reparameterization, uniform distributions on the auxiliary parameters $\boldsymbol{\eta}$ will in general imply a nonuniform distribution on the original parameters $\boldsymbol{\theta}$ (Fisher, 1930; Jaynes, 2003). Hence, it is necessary to use adjusted priors for the auxiliary parameters, which we derive in closed form for linear order constraints of the type $\theta_1 \leq \theta_2 \leq \dots \leq \theta_P$

below. Note that our results are not limited to MPT models but do also apply to other models with bounded parameter spaces (up to scaling constants).

Reparameterized Order-Constraints in a Bayesian Framework

To illustrate the importance of priors for reparameterized, order-constrained models, we use the product-binomial model, a trivial MPT model:

$$f^\theta(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^P \text{Bin}(y_i, n_i, \theta_i). \quad (3)$$

This model assumes that the frequencies y_i across P conditions are independent binomial random variables with success probabilities θ_i and number of draws n_i . In this simple model, we test the hypothesis that the P conditions are ordered with increasing success probabilities, which is equivalent to the order constraint $\theta_1 \leq \theta_2 \leq \dots \leq \theta_P$.

Based on the approach by Knapp and Batchelder (2004) outlined above, we use the auxiliary parameters $\boldsymbol{\eta}$ defined recursively by $\eta_P = \theta_P$ and $\eta_i = \theta_i / \theta_{i+1}$ representing the relative decrease from θ_{i+1} to θ_i for all $i = 1, \dots, P-1$. By replacing the original parameters, the likelihood of the reparameterized model becomes

$$f^\eta(\mathbf{y} \mid \boldsymbol{\eta}) = \prod_{i=1}^P \text{Bin}(y_i, n_i, \prod_{j=i}^P \eta_j). \quad (4)$$

Note that this is just one of many possible reparameterizations for MPT models (i.e., Method A of Knapp and Batchelder, 2004, Appendix B extends our results to Method B).

The original, order-constrained model¹ \mathcal{M}_u^θ has a uniform prior on the substantive parameters $\boldsymbol{\theta}$ and thus a constant density on the admissible parameter space,

$$\pi_u^\theta(\boldsymbol{\theta}) = P! \cdot \mathcal{I}_{\{\theta_1 \leq \dots \leq \theta_P\}}(\boldsymbol{\theta}), \quad (5)$$

where $\mathcal{I}_{\{\theta_1 \leq \dots \leq \theta_P\}}$ is the indicator function, which is one if $\theta_1 \leq \dots \leq \theta_P$ and zero otherwise. This model is henceforth termed ‘balanced model’ because the original, substantive parameter values are assigned equal prior mass.

¹We use superscripts for the parameterization and subscripts for the prior. Hence, the models \mathcal{M}_u^η and \mathcal{M}_a^η share the parameter $\boldsymbol{\eta}$ but assume a uniform and an adjusted prior (derived below), respectively.

In contrast, the reparameterized model \mathcal{M}_u^η assumes independent uniform priors on the auxiliary parameters $\boldsymbol{\eta}$ and thus a constant density,

$$\pi_u^\eta(\boldsymbol{\eta}) = \prod_{i=1}^P \mathcal{I}_{[0,1]}(\eta_i). \quad (6)$$

We call this the ‘unbalanced model,’ because –as shown in the next section– the original parameters are not equally weighted. Note that this reparameterized model can also be implemented without the explicit use of auxiliary parameters $\boldsymbol{\eta}$ by using conditional uniform priors on the original parameters, that is, assuming that the constrained parameters θ_i have a constant density on the range $[0, \theta_{i+1}]$.²

Implied Prior Distributions on the Original Parameters

The balanced model \mathcal{M}_u^θ and the unbalanced model \mathcal{M}_u^η imply different prior distributions on the parameters of interest $\boldsymbol{\theta}$ (cf. Lee and Wagenmakers, 2014, Ch. 9). To illustrate the difference, we sampled parameters from the prior distributions of both models and plotted the resulting distributions of $\boldsymbol{\theta}$. For the reparameterized model, this requires the transformation $\theta_i = \prod_{j=i}^P \eta_j$ (e.g., $\theta_1 = \eta_1 \cdot \eta_2 \cdot \eta_3$, $\theta_2 = \eta_2 \cdot \eta_3$, and $\theta_3 = \eta_3$). For $P = 2$, Panels A and B of Figure 1 show the resulting samples of $\boldsymbol{\theta}$ for each model. Whereas the balanced model puts equal probability mass on all admissible parameter values $\theta_1 \leq \theta_2$, the unbalanced model puts more mass on small values of θ_1 . Moreover, the balanced model has symmetric marginal distributions whereas the unbalanced model has a uniform distribution on the larger parameter θ_2 and overweights small values of θ_1 .

This discrepancy becomes more severe as the number P of order constraints increases. Since the differences in higher dimensions are difficult to visualize, we instead compare the univariate marginal prior distributions on θ_i . For the balanced model \mathcal{M}_u^θ , Goggans et al. (2007) showed that the marginal prior on θ_i is given by

$$\pi_u^{\theta_i}(\theta_i) \propto \theta_i^{i-1} (1 - \theta_i)^{P-i}, \quad (7)$$

which is a $\text{Beta}(i, P - i + 1)$ distribution. In contrast, the implied marginal prior on θ_i

²In the software JAGS (Plummer, 2003), this is done by `theta[i] ~ dunif(0, theta[i+1])`.

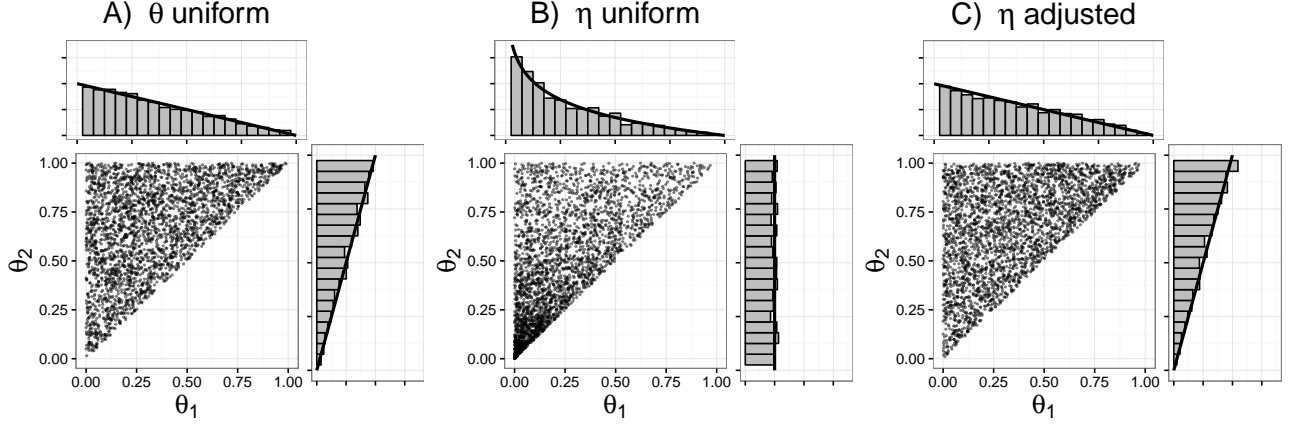


Figure 1. Assumed (Panel A) and implied (Panel B and C) prior distributions on the parameters of interest $\boldsymbol{\theta}$ for a two-dimensional model ($\theta_1 \leq \theta_2$; cf. Lee and Wagenmakers, 2014, Ch. 9).

in the unbalanced model \mathcal{M}_u^η (see Appendix A) is

$$\pi^{\eta_i}(\theta_i) = \frac{1}{(P-i)!} (-\log \theta_i)^{P-i}. \quad (8)$$

A comparison of $P = 2$ in Figure 1 and $P = 4$ in Figure 2 shows that the difference in marginal priors on $\boldsymbol{\theta}$ increases with the number of order constraints. The balanced prior model is symmetric in a sense that small values of the smallest parameter θ_1 get as much prior mass as the complementary values of the largest parameter θ_P . In contrast, the unbalanced prior model always assumes a uniform distribution on the largest parameter θ_P and overweights small values of all other parameters. Note that this asymmetry implies that any inferences based on the unbalanced model will depend on the definition of the success probabilities $\boldsymbol{\theta}$ (e.g., whether to count heads or tails when tossing a coin). Most importantly, these different priors on $\boldsymbol{\theta}$ will have a substantial impact on the marginal likelihood, since they result in a different weighting of the parameter space when averaging the likelihood (a detailed comparison can be found in the supplementary material).

Adjusted Priors for the Auxiliary Parameters

To ensure that a reparameterized model is equivalent to the original one, the priors for the auxiliary parameters need to be adjusted. For this purpose, we rewrite the

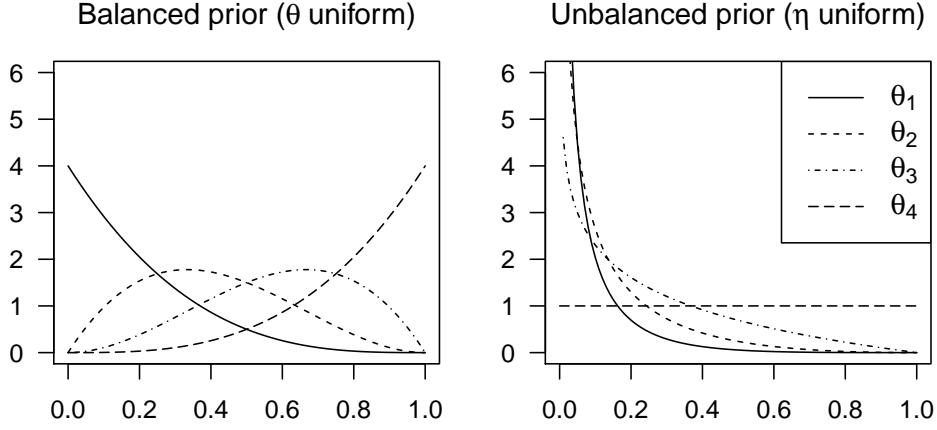


Figure 2. Marginal prior distributions on the parameters of interest $\boldsymbol{\theta}$ for a four-dimensional model ($\theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4$).

reparameterization of $\theta_1 \leq \dots \leq \theta_P$ introduced above (Method A of Knapp & Batchelder, 2004) by the one-to-one transformation function $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\theta})$,

$$\mathbf{g} : \Theta' \rightarrow [0, 1]^P \quad (9)$$

$$g_i(\boldsymbol{\theta}) = \theta_i / \theta_{i+1} \text{ for } i = 1, \dots, P-1$$

$$g_P(\boldsymbol{\theta}) = \theta_P,$$

where Θ' is the constrained parameter space, $\Theta' = \{\boldsymbol{\theta} \in [0, 1]^P : \theta_1 \leq \dots \leq \theta_P\}$. To recover the original parameters $\boldsymbol{\theta}$, the inverse of this function is required:

$$\mathbf{g}^{-1} : [0, 1]^P \rightarrow \Theta' \quad (10)$$

$$g_i^{-1}(\boldsymbol{\eta}) = \prod_{j=i}^P \eta_j \text{ for } i = 1, \dots, P.$$

In order to obtain adjusted priors for the auxiliary parameters $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\theta})$, we need to apply the changes-of-variables theorem to the marginal likelihood integral in Eq. 2,

$$\begin{aligned} p(\mathbf{y} \mid \mathcal{M}_u^\theta) &= \int_{\Theta'} f^\theta(\mathbf{y} \mid \boldsymbol{\theta}) \pi_u^\theta(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{\mathbf{g}(\Theta')} f^\theta(\mathbf{y} \mid \mathbf{g}^{-1}(\boldsymbol{\eta})) P! |\det D\mathbf{g}^{-1}(\boldsymbol{\eta})| d\boldsymbol{\eta} \\ &= \int_{[0,1]^P} f^\eta(\mathbf{y} \mid \boldsymbol{\eta}) P! |\det D\mathbf{g}^{-1}(\boldsymbol{\eta})| d\boldsymbol{\eta}, \end{aligned} \quad (11)$$

where $D\mathbf{g}^{-1}(\boldsymbol{\eta})$ is the Jacobian matrix of partial derivatives of the inverse transformation function \mathbf{g}^{-1} . Note that this requires that the transformation function \mathbf{g}

is one-to-one and continuously differentiable. For the reparameterization \mathbf{g} in Eq. 9,

$$[D\mathbf{g}^{-1}(\boldsymbol{\eta})]_{i,k} = \frac{\partial g_i^{-1}(\boldsymbol{\eta})}{\partial \eta_k} = \begin{cases} 0, & \text{if } k < i \\ 1, & \text{if } i = k = P \\ \prod_{\substack{j=i,\dots,P \\ j \neq k}} \eta_j, & \text{else} \end{cases} \quad (12)$$

For $P = 3$ parameters, this matrix is as follows:

$$\begin{pmatrix} \eta_2\eta_3 & \eta_1\eta_3 & \eta_1\eta_2 \\ 0 & \eta_3 & \eta_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Because all off-diagonal elements for $k < i$ are zero, the determinant of $D\mathbf{g}^{-1}(\boldsymbol{\eta})$ is simply given by the product of the diagonal elements, which are always positive. Hence, based on Eq. 11, it follows that the adjusted prior for the auxiliary parameters $\boldsymbol{\eta}$ is

$$\begin{aligned} \pi_a^\eta(\boldsymbol{\eta}) &= P! |\det D\mathbf{g}^{-1}(\boldsymbol{\eta})| \\ &= P! \prod_{i=1}^{P-1} \left(\prod_{j=i+1}^P \eta_j \right) \\ &= P! \prod_{i=2}^P \eta_i^{i-1} \\ &= \prod_{i=1}^P \frac{1}{B(i, 1)} \eta_i^{i-1} (1 - \eta_i)^0, \end{aligned} \quad (13)$$

which is a product of beta-distribution densities. Hence, to obtain a marginal likelihood which matches that of the balanced model with a uniform prior on the restricted parameter space, the reparameterized model requires independent beta distributions $\eta_i \sim \text{Beta}(i, 1)$ as adjusted priors for the auxiliary parameters. For the two-dimensional case $P = 2$, the different priors on $\boldsymbol{\eta}$ are illustrated in Figure 3. Obviously, the adjusted prior puts more mass on large values of η_2 compared to the uniform prior, which in return implies a uniform prior on the original parameter space (Panel C of Figure 1).

Note that this result can easily be generalized to other statistical models with bounded parameter spaces $a \leq \theta_i \leq b$ because only the support of the integral and the corresponding scaling constants in Eq. 11 change. More generally, this derivation might

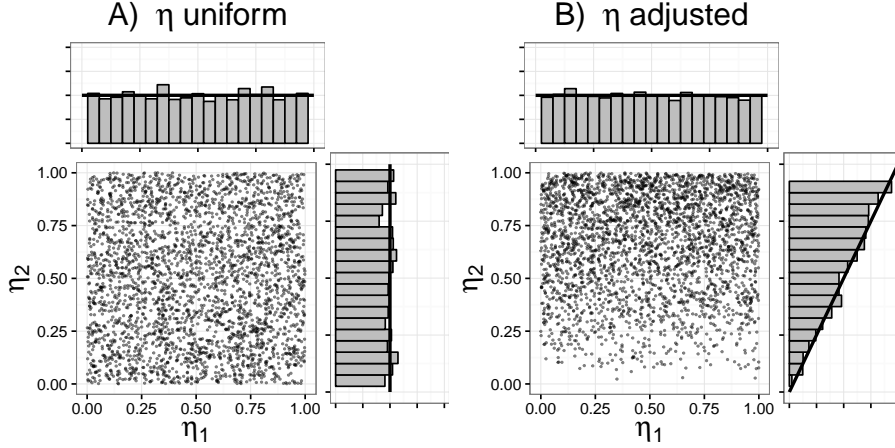


Figure 3. Prior distributions on the auxiliary parameters $\boldsymbol{\eta}$ of the reparameterized, order-constrained model $\theta_1 \leq \theta_2$ (cf. Figure 1).

serve as an example how to derive adjusted priors for any type of continuously differentiable, one-to-one reparameterization. The key insight is Eq. 11: when transforming the original parameters in the marginal likelihood integral, the changes-of-variables theorem allows to obtain an adjusted prior for the new parameters.

An Empirical Example: The Pair-Clustering Model

The following example of pair clustering in memory (Batchelder and Riefer, 1980; Batchelder and Riefer, 1986) highlights the importance of adjusted priors. The standard paradigm consists of a learning and a free-recall phase and includes words that are semantically associated (pairs) and words that are semantically unrelated (singletons). To account for the finding that pairs are usually recalled either consecutively or not at all, Batchelder and Riefer (1980) proposed the pair-clustering model shown in Figure 4. Given a pair of semantically associated words, both words are stored jointly with probability c . If storage was successful, the whole cluster is retrieved from memory with probability r resulting in consecutive recall of both words (category E_1). In contrast, if retrieval fails with probability $1 - r$, neither word is recalled (E_4). In case of unsuccessful clustering of a pair, both words can only be stored and recalled independently with probability u resulting either in non-consecutive recall of both items (E_2), in recall of one of the two items (E_3), or in recall of neither item (E_4). Similarly,

singletons can only be stored and retrieved separately with probability u resulting either in recall (F_1) or not (F_2).

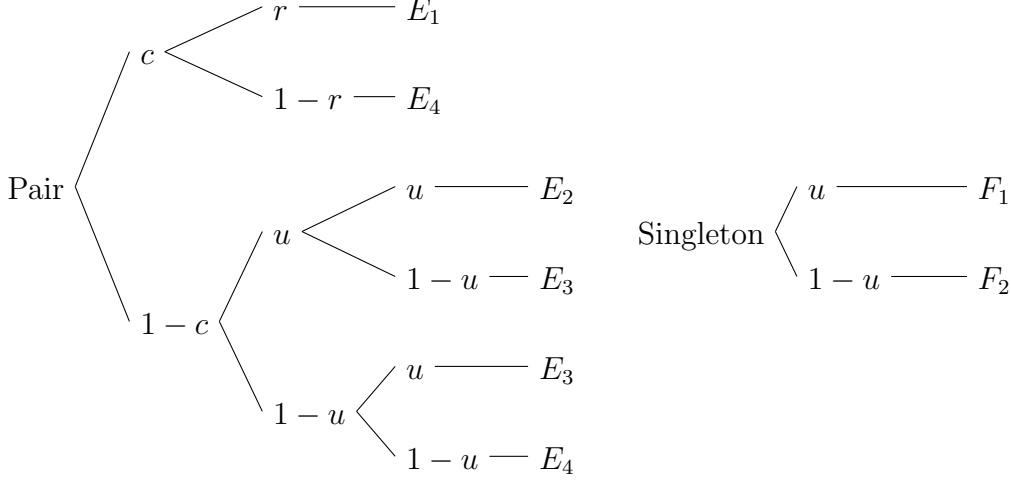


Figure 4. The pair-clustering model (Batchelder and Riefer, 1986). See text for details.

Riefer et al. (2002) used the pair-clustering model to test the specificity of memory deficits in schizophrenic patients. For this purpose, six study-test trials were administered to compare the memory performance of healthy controls and schizophrenics. Riefer et al. (2002) predicted that the three memory parameters for storage (c), retrieval (r), and single-item memory (u) should increase over trials due to learning. For the storage parameter c , this hypothesis is represented by the order constraints $c_{i1} \leq \dots \leq c_{i6}$ within each group $i = 1, 2$ (the same order constraints hold for r_{ik} and u_{ik}). In line with this prediction, the posterior estimates, shown in Figure 5, indeed indicate an increase of memory performance across trials in both groups.

We used the importance sampling approach by Vandekerckhove, Matzke, and Wagenmakers (2015)³ to compute the marginal likelihoods of four models: the balanced, order-constrained model \mathcal{M}_a^η with a uniform prior on the original parameter space (reparameterized with adjusted priors); the unbalanced, reparameterized model \mathcal{M}_u^η with uniform auxiliary parameters; the full model \mathcal{M}_f^θ without constraints on c_{ij} , r_{ij} , and u_{ij} ; and the null model \mathcal{M}_n^θ with only three parameters c_i , r_i , and u_i that are

³See the supplementary material for a detailed explanation and implementation in R.

constant across study-test trials for each group $i = 1, 2$.

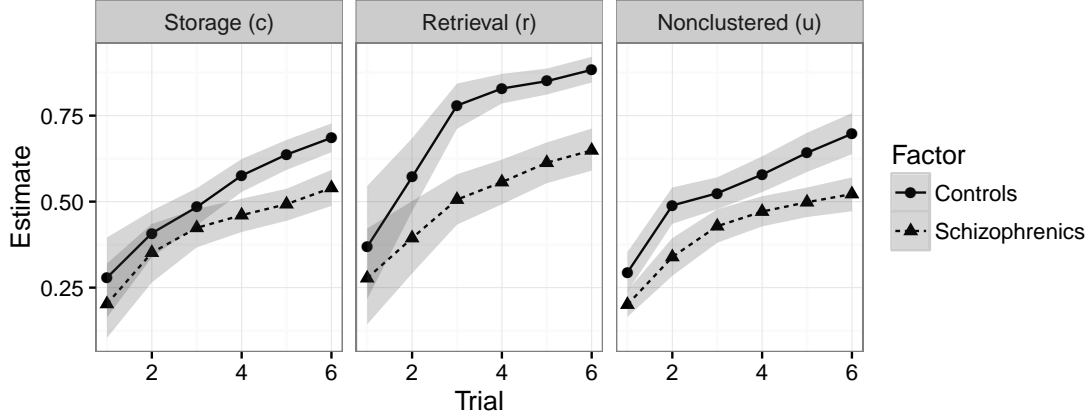


Figure 5. Mean posterior estimates and 95% credibility intervals for Experiment 3 of Riefer et al. (2002) based on the order-constrained model \mathcal{M}_a^η with adjusted priors on the auxiliary parameters.

Table 1 shows that the balanced model \mathcal{M}_a^η performed best by several orders of magnitude. Importantly, the difference in marginal likelihoods between the models \mathcal{M}_a^η and \mathcal{M}_u^η was considerable as indicated by a log Bayes factor of 21.06, or equivalently, a Bayes factor of around 1.4 billion. This huge discrepancy is noteworthy because both models implement the same order constraints and differ only in the prior distribution on the restricted parameter space. Obviously, there is much more evidence for a uniform distribution on the order-constrained, original parameter space as represented by the model \mathcal{M}_a^η . The Bayes factor for the two order-constrained models is even larger than that comparing the reparameterized model \mathcal{M}_a^η against the full model \mathcal{M}_f^θ . Hence, in terms of evidence, the choice of priors for the auxiliary parameters weighs stronger than the order constraint itself.

Discussion

Order constraints are important for testing psychological theories and are often implemented by reparameterizing the original parameters, especially in MPT modeling (Knapp and Batchelder, 2004). However, whereas inferences based on maximum likelihood are not affected by reparameterization, care has to be taken when relying on

Table 1

Estimated marginal likelihoods for Experiment 3 of Riefer et al. (2002).

Model \mathcal{M}	$p(\mathbf{y} \mid \mathcal{M})$	SE_p	$\log p(\mathbf{y} \mid \mathcal{M})$	$\text{SE}_{\log p}$
Full	7.26×10^{-87}	1.45×10^{-87}	-198.68	0.10
Balanced	3.30×10^{-74}	2.55×10^{-75}	-169.30	0.06
Unbalanced	2.64×10^{-83}	3.58×10^{-84}	-190.36	0.08
Null	6.98×10^{-264}	1.06×10^{-266}	-605.94	< 0.01

Bayesian model selection. Using a simple product-binomial model, we illustrated the difference between a uniform prior on the order-constrained, original parameters versus a uniform prior on the auxiliary parameters of the reparameterized model. Whereas the former puts equal prior mass to all of the admissible original parameters, the latter puts more weight on small parameter values (Figure 1). We derived adjusted priors for the auxiliary parameters (beta distributions) that result in identical marginal likelihoods as those obtained from the original model. In practice, the Bayes factor can differ substantially if the prior does not take the reparameterization into account as illustrated in the pair-clustering example (Riefer et al., 2002).

The impact of priors shown in the present paper might support previous criticism of the Bayes factor due to its prior sensitivity (e.g., Liu and Aitkin, 2008). However, the prior sensitivity can be seen as one of the major advantages of Bayesian model selection (e.g., Vanpaemel, 2010). By incorporating prior knowledge into a cognitive model, the model becomes more specific and makes stronger predictions. In line with this view, order constraints can be interpreted as highly informative priors that put zero probability on large proportions of the original parameter space (Lee, 2016).

Importantly, constraining the prior probability to be zero is qualitatively different from assuming very low but nonzero probability mass. If the prior is zero on some parameter subspace, the posterior will always be zero on this subspace regardless of the data. In contrast, for an arbitrarily small but nonzero prior probability, it is in principle possible to observe data that substantially increase the posterior probability on this subspace.

Therefore, if the Bayes factor is criticized for its sensitivity with respect to the prior distribution, the same critique applies to order constraints in general, which merely encode highly informative prior knowledge about the parameters.

Conclusion

The Bayes factor has many advantages that are desirable for model selection (Wagenmakers, 2007) — it directly emerges from the application of Bayes’ theorem when computing posterior model odds, it has a direct interpretation as the evidence in favor of one model versus another, and it considers order constraints and the functional complexity of statistical models (Myung and Pitt, 1997). However, if order constraints are reparameterized, priors on the new auxiliary parameters need to be adjusted to ensure meaningful priors on the original parameters of interest. For linear order constraints in MPT models, we showed that these adjusted priors are independent beta distributions. With the present paper, we hope to raise awareness about this issue and show the severe consequences of using default priors for reparameterized models.

References

- Batchelder, W. H. and D. M. Riefer (1980). “Separation of storage and retrieval factors in free recall of clusterable pairs.” In: *Psychological Review* 87, pp. 375–397. DOI: 10.1037/0033-295X.87.4.375.
- (1986). “The statistical analysis of a model for storage and retrieval processes in human memory.” In: *British Journal of Mathematical and Statistical Psychology* 39, pp. 129–149. DOI: 10.1111/j.2044-8317.1986.tb00852.x.
- (1999). “Theoretical and empirical review of multinomial process tree modeling.” In: *Psychonomic Bulletin & Review* 6, pp. 57–86. DOI: 10.3758/BF03210812.
- Bröder, Arndt et al. (2008). “Different storage and retrieval deficits in normal aging and mild cognitive impairment: A multinomial modeling analysis.” In: *Psychology and Aging* 23, pp. 353–365. DOI: 10.1037/0882-7974.23.2.353.
- Erdfelder, Edgar et al. (2009). “Multinomial processing tree models: A review of the literature.” In: *Journal of Psychology* 217, pp. 108–124. DOI: 10.1027/0044-3409.217.3.108.
- Fisher, R. A. (1930). “Inverse Probability.” In: *Proceedings of the Cambridge Philosophical Society* 26, pp. 528–535. DOI: 10.1017/S0305004100016297.
- Goggans, Paul M. et al. (2007). “Assigning priors for ordered and bounded parameters.” In: *AIP Conference Proceedings*. Vol. 954. AIP Publishing, pp. 276–282. DOI: 10.1063/1.2821273.
- Heck, D. W. and E Erdfelder (in press). “Extending multinomial processing tree models to measure the relative speed of cognitive processes.” In: *Psychonomic Bulletin & Review*.
- Hilbig, Benjamin E. and Morten Moshagen (2014). “Generalized outcome-based strategy classification: Comparing deterministic and probabilistic choice models.” In: *Psychonomic Bulletin & Review* 21, pp. 1431–1443. DOI: 10.3758/s13423-014-0643-0.
- Hojtink, H. (2011). *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Boca Raton, FL: Chapman & Hall/CRC.

- Iverson, Geoffrey J. (2006). “An essay on inequalities and order-restricted inference.” In: *Journal of Mathematical Psychology* 50, pp. 215–219. DOI: 10.1016/j.jmp.2006.01.007.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
- Jeffreys, H (1961). *Theory of probability*. 3rd. New York: Oxford University Press.
- Kass, Robert E and Adrian E Raftery (1995). “Bayes factors.” In: *Journal of the American Statistical Association* 90, pp. 773–795. DOI: 10.1080/01621459.1995.10476572.
- Kellen, D. and K. C. Klauer (2015). “Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions.” In: *Psychological Review* 122, pp. 542–557.
- Klauer, K. C. (2010). “Hierarchical multinomial processing tree models: A latent-trait approach.” In: *Psychometrika* 75, pp. 70–98. DOI: 10.1007/s11336-009-9141-0.
- Klauer, K. C., H. Singmann, and D. Kellen (2015). “Parametric order constraints in multinomial processing tree models: An extension of Knapp and Batchelder (2004).” In: *Journal of Mathematical Psychology* 64, pp. 1–7. DOI: 10.1016/j.jmp.2014.11.001.
- Klugkist, Irene and Herbert Hoijtink (2007). “The Bayes factor for inequality and about equality constrained models.” In: *Computational Statistics & Data Analysis* 51, pp. 6367–6379. DOI: 10.1016/j.csda.2007.01.024.
- Klugkist, Irene, Olav Laudy, and Herbert Hoijtink (2005). “Inequality constrained analysis of variance: A Bayesian approach.” In: *Psychological Methods* 10, p. 477.
- Knapp, Bethany R. and William H. Batchelder (2004). “Representing parametric order constraints in multi-trial applications of multinomial processing tree models.” In: *Journal of Mathematical Psychology* 48, pp. 215–229. DOI: 10.1016/j.jmp.2004.03.002.
- Lee, Michael D. (2016). “Bayesian outcome-based strategy classification.” In: *Behavior Research Methods* 48, pp. 29–41. DOI: 10.3758/s13428-014-0557-9.

- Lee, Michael D. and Eric-Jan Wagenmakers (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Liu, Charles C. and Murray Aitkin (2008). “Bayes factors: Prior sensitivity and model generalizability.” In: *Journal of Mathematical Psychology* 52, pp. 362–375. DOI: 10.1016/j.jmp.2008.03.002.
- Moshagen, M. (2010). “multiTree: A computer program for the analysis of multinomial processing tree models.” In: *Behavior Research Methods* 42, pp. 42–54. DOI: 10.3758/BRM.42.1.42.
- Myung, In Jae and Mark A. Pitt (1997). “Applying Occam’s razor in modeling cognition: A Bayesian approach.” In: *Psychonomic Bulletin & Review* 4, pp. 79–95.
- Myung, J. I. (2000). “The importance of complexity in model selection.” In: *Journal of Mathematical Psychology* 44, pp. 190–204. DOI: 10.1006/jmps.1999.1283.
- Myung, Jay I., George Karabatsos, and Geoffrey J. Iverson (2005). “A Bayesian approach to testing decision making axioms.” In: *Journal of Mathematical Psychology* 49, pp. 205–225. DOI: 10.1016/j.jmp.2005.02.004.
- Plummer, Martyn (2003). “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.” In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vol. 124. Vienna, Austria, p. 125.
- Riefer, David M. et al. (2002). “Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models.” In: *Psychological Assessment* 14, pp. 184–201. DOI: 10.1037/1040-3590.14.2.184.
- Singmann, Henrik and David Kellen (24, 2013). “MPTinR: Analysis of multinomial processing tree models in R.” In: *Behavior Research Methods* 45, pp. 560–575. DOI: 10.3758/s13428-012-0259-0.
- Smith, J. B. and W. H. Batchelder (2010). “Beta-MPT: Multinomial processing tree models for addressing individual differences.” In: *Journal of Mathematical Psychology* 54, pp. 167–183. DOI: 10.1016/j.jmp.2009.06.007.

Vandekerckhove, J. S., D. Matzke, and E. Wagenmakers (2015). “Model comparison and the principle of parsimony.” In: *Oxford Handbook of Computational and Mathematical Psychology*. New York, NY: Oxford University Press, pp. 300–319.

Vanpaemel, Wolf (2009). “Measuring model complexity with the prior predictive.” In: *Advances in Neural Information Processing Systems 22*. Ed. by Y. Bengio et al. Curran Associates, Inc., pp. 1919–1927.

— (2010). “Prior sensitivity in theory testing: An apologia for the Bayes factor.” In: *Journal of Mathematical Psychology* 54, pp. 491–498. DOI: 10.1016/j.jmp.2010.07.003.

Wagenmakers, Eric-Jan (2007). “A practical solution to the pervasive problems of p values.” In: *Psychonomic Bulletin & Review* 14, pp. 779–804. DOI: 10.3758/BF03194105.

Wagenmakers, Eric-Jan et al. (2015). “A power fallacy.” In: *Behavior Research Methods* 47, pp. 913–917. DOI: 10.3758/s13428-014-0517-4.

Appendix A

Marginal Priors for the Reparameterized Model

We derive the marginal priors π^{θ_i} on the original parameters θ_i implied by the unbalanced prior model with uniform prior distributions for the auxiliary parameters $\boldsymbol{\eta}$. More precisely, given the prior distribution

$$\pi_u^\eta(\boldsymbol{\eta}) = \prod_{i=1}^P \mathcal{I}_{[0,1]}(\eta_i), \quad (14)$$

it follows that the original parameters $\theta_i = \prod_{j=i}^P \eta_j$ have the marginal distribution

$$\pi^{\theta_i}(\theta_i) = \frac{1}{(P-i)!} (-\log \theta_i)^{P-i} \quad (15)$$

for $\theta_i \in [0, 1]$. We prove this by backward induction. Obviously, $\theta_P = \eta_P \in [0, 1]$ and hence

$$\pi^{\theta_P}(\theta_P) = \mathcal{I}_{[0,1]}(\theta_P) = \frac{1}{(P-P)!} (-\log \theta_P)^{P-P}.$$

Next, assume that Eq. 15 holds for $i \in \{2, \dots, P\}$. Then, for $i \rightarrow i-1$, the cumulative density function of θ_{i-1} is given by

$$\begin{aligned} F^{\theta_{i-1}}(x) &= \Pr[\eta_{i-1}\theta_i \leq x] \\ &= \int_0^1 \Pr[\eta_{i-1} \leq \frac{x}{t}] \pi_r^{\theta_i}(t) dt \\ &= \frac{1}{(P-i)!} \left[\int_0^x 1 \cdot (-\log t)^{P-i} dt + x \int_x^1 \frac{1}{t} (-\log t)^{P-i} dt \right] \end{aligned}$$

Taking the derivative d/dx of the cumulative density function results in

$$\begin{aligned} \pi^{\theta_{i-1}}(x) &= \frac{1}{(P-i)!} \left[(-\log x)^{P-i} + \int_x^1 \frac{1}{t} (-\log t)^{P-i} dt - x \cdot \frac{1}{x} (-\log x)^{P-i} \right] \\ &= \frac{1}{(P-i)!} \int_x^1 \frac{1}{t} (-\log t)^{P-i} dt \\ &= \frac{1}{(P-(i-1))!} (-\log x)^{P-(i-1)}, \end{aligned}$$

which completes the proof.

Appendix B

Adjusted Priors for Method B of Knapp & Batchelder (2004)

Method B.

Whereas the auxiliary parameters $\boldsymbol{\eta}$ in the main text are defined as the proportional decreases of the original parameters $\theta_P, \theta_{P-1}, \dots, \theta_1$, order constraints can also be defined in terms of proportional increases of $\theta_1, \theta_2, \dots, \theta_P$. In this approach, called Method B by Knapp and Batchelder (2004), the continuously differentiable, one-to-one transformation function $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\theta})$ is given by

$$\begin{cases} g_1(\boldsymbol{\theta}) = \theta_1 \\ g_i(\boldsymbol{\theta}) = \frac{\theta_i - \theta_{i-1}}{1 - \theta_{i-1}}, \end{cases}$$

in which the parameters η_i can be interpreted as ‘growth parameters.’ The inverse is given by

$$g_i^{-1}(\boldsymbol{\eta}) = 1 - \prod_{j=1}^i (1 - \eta_j), \quad (16)$$

Moreover, the Jacobian $D\mathbf{g}^{-1}(\boldsymbol{\eta})$ is

$$[D\mathbf{g}^{-1}(\boldsymbol{\eta})]_{i,k} = \frac{\partial g_i^{-1}(\boldsymbol{\eta})}{\partial \eta_k} = \begin{cases} 0, & \text{if } k < i \\ 1, & \text{if } i = k = 1 \\ \prod_{\substack{j=1, \dots, i \\ j \neq k}} (1 - \eta_j), & \text{else.} \end{cases} \quad (17)$$

The determinant of this triangular matrix is the product of the diagonal elements, which leads to the adjusted prior

$$\begin{aligned} \pi_a^\eta(\boldsymbol{\eta}) &= P! |\det D\mathbf{g}^{-1}(\boldsymbol{\eta})| \\ &= P! \prod_{i=2}^P \left(\prod_{j=1}^{i-1} (1 - \eta_j) \right) \\ &= P! \prod_{i=1}^{P-1} (1 - \eta_i)^{P-i} \\ &= \prod_{i=1}^P \frac{1}{B(1, P - i + 1)} \eta_i^0 (1 - \eta_i)^{P-i}. \end{aligned} \quad (18)$$

It follows that the adjusted prior for the reparameterized model requires independent beta priors on the auxiliary parameters, that is, $\eta_i \sim \text{Beta}(1, P - i + 1)$.